BU-1005-M

INTERPRETING BLOCKS AND RANDOM FACTORS

by

Myra L. Samuels, George Casella and George P. McCabe
Purdue University and Cornell University *

Department of Statistics
Purdue University

July 1990

* This research was performed while George Casella was Visiting Professor, Department of Statistics, Purdue University. George Casella's work was partially supported by NSF Grant #DMS89–0039.

*

INTERPRETING BLOCKS AND RANDOM FACTORS

by

Myra L. Samuels, George Casella and George P. McCabe
Purdue University and Cornell University

Abstract

Can blocks be tested in a randomized blocks design? It is well-known that two different formulations of the linear mixed model yield conflicting answers to this question. This paper examines the model formulations from the point of view of statistical relevance. It is found that the question of testing blocks is not the same as the question of testing a random factor. Viewing blocking as a device to increase efficiency leads to a hypothesis concerning blocks which is distinct from the hypothesis of no "block main effect." Tests of the two hypotheses, and interval estimation of the impact of blocking on efficiency, are described. The merits of the two model formulations are compared, and recommendations are made to enhance the clarity and heuristic usefulness of mixed linear models.

KEY WORDS: Randomized blocks, Mixed model; Random factors; Analysis of variance; Linear models; Expected mean squares; Variance Components; Interaction; Best Linear Unbiased Prediction.

1. Introduction

In a two–way mixed model, how should one test the random factor if additivity is not assumed? Can blocks be tested in a randomized blocks design if blocks are regarded as random and additivity is not assumed? Although these questions have been around for several decades (Cornfield and Tukey, 1956; Wilk and Kempthorne, 1955, 1956), there is still some confusion as to the answers. The conflict is dramatically illustrated by the fact that the widely used SAS statistical package produces tables of expected mean squares (EMSs) which disagree with those given in many textbooks, and with another widely used statistical package, BMDP.

This paper will address these questions, with special attention to the important case of blocked designs. We consider not only randomized blocks designs but also nonrandomized observational designs which incorporate blocking.

To set the stage, consider the following examples of blocked designs. (Although the examples all involve human beings, our discussion is meant to be quite general.)

Example 1.1. Pairs of female twins are randomly selected from a source population to participate in a study of an anti–aging skin cream. One randomly chosen member of each pair receives the cream (Treatment 1) and the other receives placebo (Treatment 2). The observation Y is a measure of average skin thickness on the two forearms.

Example 1.2. The treatments are as in Example 1.1, but now the subjects are women (not twins) randomly selected from a source population. The two treatments are randomly allocated to the two arms of each subject, and skin thickness Y is measured on each arm.

Example 1.3. As in Example 1.2, the subjects are women randomly selected from a source population. The question of interest now is whether the density of neurons is greater in the right or the left hemisphere of the brain. Each measurement Y of neural density is taken from a CAT scan of the head.

Example 1.4. As in Example 1.1, pairs of female twins are observed, but now the twins are selected from a source population in which one member of the pair smokes cigarettes

and the other does not. The observation Y is a measure of cardiac health.

Examples 1.1 and 1.2 are commonly called randomized blocks designs. In Example 1.1 the block is a pair of subjects, while in Example 1.2 each subject is a block. In the psychometric literature, Example 1.2 is also called a repeated measures design.

Examples 1.3 and 1.4 are blocked designs but are strictly observational: no randomized allocation is involved. Note that the primary purpose of blocking in Example 1.4 is to control confounding – that is, to remove the effect of covariates (such as age and genetic background) which might otherwise distort the observed relationship between smoking and cardiac health. By contrast, in Examples 1.1, 1.2 and 1.3 a valid study could be conducted without blocking, but blocking is expected to increase efficiency.

For designs such as the above, a standard statistical approach – a mixed–model analysis of variance (ANOVA) – leads to two conflicting answers as to whether the ANOVA table yields an F test for blocks. More generally, there are two conflicting answers as to which mean square should form the denominator of an F statistic for testing the random factor in a two–way mixed model. Several authors (Hartley and Searle 1969; Searle 1971a, Searle 1971b, pp. 401–404; Hocking 1973; Harville 1978; Hocking 1985, pp. 330-334) have discussed the conflict and have noted that in a certain sense the discrepancy can be resolved by suitable re–definition of the model parameters. This is an algebraic resolution, but it is not a statistical resolution because it sidesteps the question of which parameterization is preferable. We will discuss the problem from a statistical point of view and try to develop some statistical perspective to aid the experimenter who asks: Can I test blocks? Should I? And, if so, how?

In Section 2 the two conflicting answers are exhibited. In Section 3 it is argued that the effect of blocks, interpreted as the impact of blocking on the efficiency of treatment comparisons, can be both tested and estimated. In Section 4 two mixed models are defined, and the paradox of the conflicting ANOVA tables is resolved by noting that there are two distinct hypotheses concerning blocks, with different statistical interpretations. In Section 5 the discussion is extended to the case of replicated measurements. In Section 6 further

comparisons of the two mixed models are considered. In Section 7 the statistical resolution of the conflict is summarized and discussed, and definitions are proposed which are intended to reduce confusion in the formulation of linear mixed models.

We assume throughout that the covariance matrix of the observations is compound symmetric. (A more general model, in which compound symmetry is not assumed, is discussed in Samuels, Casella and McCabe 1990.) In discussing formal inference, we will assume that the random variables in the models are jointly normal. Much of our development, however, involves only first and second moments and is not dependent on normality assumptions.

2. The Issue

For simplicity, we will use the term "treatments" throughout our discussion, even though this usage is unnatural for an application like Example 1.3.

Let the data be represented by Y_{ijk}, where $i = 1, \ldots, I$ represents treatments, $j = 1, \ldots, J$ represents blocks and $k = 1, \ldots, K$ represents repeated observations on the same treatment–block combination. (Although $K = 1$ in Examples 1.1—1.4 and in many blocked designs, we will find it helpful to also consider the case $K > 1$.) Table 2.1 shows the usual orthogonal decomposition of the total sum of squares, including terms for treatments, blocks, the treatment by block interaction, and variation within a treatment–block combination. (An overbar and a dot denote averaging over a subscript.)

Table 2.1 goes here

Corresponding to the decomposition of the sum of squares, it is natural to think in terms of an equation of the form

$$Y_{ijk} = \mu + T_i + B_j + G_{ij} + e_{(ij)k} \tag{2.1}$$

where μ represents an overall mean, T_i the treatment effect, B_j the block effect, G_{ij} the treatment–block interaction and $e_{(ij)k}$ the residual variation. (The bracketed subscript denotes nesting of the "e" effect within a treatment–block combination.) The equation (2.1) becomes a statistical model useful for guiding the analysis of data only when constraints and assumptions regarding the distributions of the various terms are specified. These issues will be addressed in detail in Section 4.

For general analysis of variance models, construction of F tests or quasi–F tests for testing hypotheses under normality assumptions is commonly based on a table of EMSs. The controversy regarding blocks arises because, using models that appear similar on the surface, it is possible to calculate different expressions for the EMSs. The case where disagreement exists is the mixed model, in which treatments are regarded as fixed and blocks are regarded as random.

Routine application of the "EMS algorithm" (see, for example, Winer 1971, Hicks 1982, or Kirk 1982) to (2.1) with treatments fixed and the other terms random gives the first column (Version 1) of Table 2.2. Examination of these EMSs suggests that the denominator for testing blocks should be MS(Within T*B). In the important case where $K = 1$, this term has no degrees of freedom, which seems to indicate that blocks cannot be tested. The Version 1 table (or a slight variant) is presented in many textbooks (for instance, Scheffe, 1959; Searle, 1971, p. 403; Winer, 1971; Steel and Torrie, 1980; Hicks, 1982; Kirk, 1982; Montgomery, 1984; Snedecor and Cochran, 1989) and by the computer program BMDP8V.

Table 2.2 goes here

On the other hand, the EMSs presented in the second column (Version 2) of Table 2.2 are also found in textbooks (for instance, Searle, 1971, p. 401) and are given by the RANDOM statement in the SAS procedure GLM. This EMS table can be obtained via

the EMS algorithm from the following slight modification of (2.1):

$$Y_{ijk\ell} = \mu + T_i + B_j + G_{\ell(ij)} + e_{(ij\ell)k}$$

where $\ell \equiv 1$. The Version 2 EMSs suggest that the denominator for testing blocks should be MS(T*B), and, in particular, that blocks can be tested when $K = 1$.

The EMS algorithm can be a convenient tool to guide an analysis, but we see that it can give different answers when applied to what appear to be very similar representations for the same set of data. Furthermore, statistical practice should not be determined by which software package is available to the user. We will try to clarify the situation by translating (2.1) into more fully specified models.

In the models we consider, the random variables are generated by sampling from (infinite) populations. Before proceeding, we note that other models are sometimes used for a randomized blocks analysis. For example, one might use the restriction error models of Anderson (Anderson 1970, Anderson and McLean 1974) or randomization (permutation) models (Kempthorne 1952, pp. 135–151, White 1975). Such models are based on different assumptions than the models considered here, and can lead to different answers. The models we consider, sometimes called population models, represent a common approach to modeling a statistical analysis. (Note that randomization models would be unnatural for Examples 1.3 and 1.4, since those examples do not involve randomized allocation.)

Recently, Lentner, Arnold and Hinkelmann (1989) have discussed the randomized blocks design. They assume treatment-block additivity, and are concerned with the question of how to assess the effect of blocking on efficiency; they give an expression for estimated relative efficiency which differs only slightly from the one we give in Section 3. The arguments of Lentner, *et al*, are framed in terms of a randomization model, which leads them to assert that blocks cannot be tested at all, even when treatment-block additivity is assumed. By contrast, in the population model that we adopt, blocks can unquestionably be be tested when additivity is assumed; our concern is to untangle the special considerations which arise when interaction may be present.

In the next two sections we assume that $K = 1$, that is, there is only one observation

per treatment–block combination. This simplifies the notation, and it will become clear in Section 5 that our discussion extends immediately to the case $K > 1$.

3. To Block or Not to Block

In this section the testing and estimation of block effects will be approached from first principles, rather than from the viewpoint of a linear model like (2.1). In Section 4 the results will be related to two different versions of (2.1).

When an investigator considers the hypothesis that "There is no effect due to blocks," what exactly is meant? One reasonable interpretation is that the study could just as well have been done without blocks, that is, as a completely randomized design (e.g., in Examples 1.1 and 1.2), or with independent sampling rather than paired sampling (e.g., in Examples 1.3 and 1.4). In other words, the hypothesis is interpreted to mean that nothing was accomplished by blocking.

What is blocking intended to accomplish? The answer can vary according to the nature of the study. When blocking serves the purpose of controlling confounding (as in Example 1.4), the conditions under which blocks can be ignored are complex (see, for instance, Samuels 1981). We will not consider this case, but rather will limit our attention to cases (as in Examples 1.1—1.3) where the purpose of blocking is to increase the efficiency of treatment comparisons. Thus, we consider only the situation in which a non–blocked design would be a valid alternative to the blocked design.

3.1 A General Model

For the case $K = 1$ we suppress the third subscript and let

$$\{(Y_{1j}, Y_{2j}, \ldots, Y_{Ij})'; \; j = 1, \ldots, J\}$$

be independently and identically distributed (iid) random vectors each distributed as the vector $(Y_1, Y_2, \ldots, Y_I)'$. (Note that this formulation implicitly regards blocks as a random factor, for otherwise the random vectors would not be identically distributed.)

We assume that the covariance matrix of $(Y_1, \ldots, Y_I)'$ is compound symmetric — that

is, the Y_i have the same variance and the pairwise correlations are the same:

$$\begin{aligned} \mathrm{Var}(Y_i) &= \sigma_Y^2 \\ \mathrm{Cov}(Y_i, Y_{i'}) &= \rho\sigma_Y^2 \end{aligned} \tag{3.1}$$

When necessary for deriving confidence intervals and tests, we will also assume that $(Y_1, \ldots, Y_I)'$ is multivariate normal.

The parameters of central interest are the treatment means $E(Y_i)$; comparisons among these can be expressed in terms of contrasts of the form $\Sigma c_i E(Y_i)$, for some constants c_i with $\Sigma c_i = 0$. The precision with which a contrast can be estimated depends upon the variance of $(\Sigma c_i Y_i)$; thus, to compare a blocked design with a non–blocked design it is appropriate to compare the variances of $(\Sigma c_i Y_i)$ under the two designs.

For a non–blocked design, the Y_i are modeled as independent random variables, so that

$$\mathrm{Var}_N\left(\sum_i c_i Y_i\right) = \sigma_Y^2 \sum_i c_i^2$$

whereas for the blocked design we have

$$\begin{aligned} \mathrm{Var}_B\left(\sum_i c_i Y_i\right) &= \sum_i c_i^2 \sigma_Y^2 + \sum_{i \neq i'} \sum c_i c_{i'} \rho \sigma_Y^2 \\ &= \sigma_Y^2 \sum_i c_i^2 + \rho\sigma_Y^2 \left[\left(\sum_i c_i\right)^2 - \sum_i c_i^2\right] \\ &= \sigma_Y^2 (1-\rho) \sum_i c_i^2 \end{aligned}$$

because $\sum_i c_i = 0$. Thus, the ratio λ of the variances is

$$\lambda = \frac{\mathrm{Var}_N(\Sigma c_i Y_i)}{\mathrm{Var}_B(\Sigma c_i Y_i)} = \frac{1}{1-\rho}. \tag{3.2}$$

Note that λ is invariant with respect to the choice of the $\{c_i\}$. The parameter λ is a natural measure of the effect of blocking; it is the relative efficiency* of the two designs. To

* Strictly speaking, (3.2) is the asymptotic relative efficiency, since it does not account for the loss in degrees of freedom due to blocking.

obtain the same variance as a blocked design with J observations on each treatment, λJ observations on each treatment would be required in a non-blocked design. Note that if ρ is negative, then blocking results in a loss, rather than a gain, of efficiency.

From Equation (3.2) it is clear that in the context of this general model, questions regarding the effect of blocking can be addressed by consideration of the parameter ρ. In particular, the hypothesis of no block effect is expressed as

$$H_0: \rho = 0$$

and a measure of the effect of blocking is given by an estimate of ρ.

3.2 Testing for the Effect of Blocking on Efficiency

We consider first the case of two treatments, and, assuming normality of (Y_1, Y_2), derive the F–test of $H_0: \rho = 0$ from first principles. Table 3.1 gives, for $I = 2$, the ANOVA table with expected mean squares derived from the general formulation of Section 3.1. From Table 3.1 it follows that

$$\begin{aligned} \text{EMS(Blocks)} &= \tfrac{1}{2}\text{Var}(Y_1 + Y_2) = \sigma_Y^2(1+\rho) \\ \text{EMS(T*B)} &= \tfrac{1}{2}\text{Var}(Y_1 - Y_2) = \sigma_Y^2(1-\rho) \end{aligned} \tag{3.3}$$

Thus, under the hypothesis $H_0: \rho = 0$, MS(Blocks) and MS(T*B) have the same expectation.

Table 3.1 goes here

If we now assume that the distribution of $(Y_1, Y_2)'$ is normal, then MS(Blocks) and MS(T*B) are each distributed as a scaled chi-squared random variable. Then the facts that

$$\text{Cov}(Y_1 + Y_2,\ Y_1 - Y_2) = \text{Var}(Y_1) - \text{Var}(Y_2) = 0 \tag{3.4}$$

and that MS(Blocks) is a function only of $(Y_1 + Y_2)$ and MS(T*B) is a function only of $(Y_1 - Y_2)$, together with normality, imply that MS(Blocks) and MS(T*B) are independent. Therefore, under the hypothesis $H_0: \rho = 0$ we have

$$\frac{\text{MS(Blocks)}}{\text{MS(T*B)}} \sim F_{J-1,J-1}, \tag{3.5}$$

where $F_{m,n}$ is an F random variable with m and n degrees of freedom.

For the case of I treatments, an extension of the preceding argument establishes that, under the hypothesis $H_0: \rho = 0$ and assuming normality, we have

$$\frac{\text{MS(Blocks)}}{\text{MS(T*B)}} \sim F_{J-1,(I-1)(J-1)}. \tag{3.6}$$

Referring back to Table 2.2, note that the F test for blocks based on (3.6) is suggested by the Version 2 EMSs but not by the Version 1 EMSs. We will return to this point in Section 4.

3.3 Estimation of the Effect of Blocking on Efficiency

In addition to testing, an investigator may also be interested in estimating the magnitude of the increase (or, perhaps, decrease) in efficiency due to blocking. Such an estimate would be helpful in planning future studies in similar settings. For example, if blocking is costly or inconvenient, and the anticipated gain in efficiency is small, then the investigator might opt for a completely randomized design.

From equation (3.2) we saw that the relative efficiency λ is simply related to the parameter ρ. For a researcher interested in estimating the relative efficiency, an estimate of ρ is therefore needed.

For any i and i', $i \neq i'$, we can write ρ as

$$\rho = \frac{\text{Cov}(Y_i, Y_{i'})}{\text{Var}(Y_i)}. \tag{3.7}$$

In view of (3.7), a natural estimator of ρ is

$$\tilde{r} = \frac{\sum_{i \neq i'} \sum_j (Y_{ij} - \overline{Y}_{i\cdot})(Y_{i'j} - \overline{Y}_{i'\cdot})}{(I-1) \sum_i \sum_j (Y_{ij} - \overline{Y}_{i\cdot})^2},$$

which can be interpreted as the average observed covariance divided by the average observed variance. In fact, $\tilde{r}$ is the maximum likelihood estimate (MLE) of ρ, assuming normality and the constraints (3.1) (Kristof 1963, Mehta and Gurland 1969). The MLE of λ is $1/(1-\tilde{r})$.

It is easy to show that $\tilde{r}$ can be written in terms of the ANOVA mean squares as (see Hocking 1985, p. 325)

$$\tilde{r} = \frac{\text{MS(Blocks)} - \text{MS(T*B)}}{\text{MS(Blocks)} + (I-1)\text{MS(T*B)}}.$$

The statistic $\tilde{r}$ is sometimes called an intraclass correlation coefficient or a reliability coefficient; more often, however, these names are given to a somewhat different statistic (see Winer 1971, pp. 286–287; Snedecor and Cochran 1989, p. 243).

We turn now to the problem of setting confidence limits on ρ, and thereby on λ. Assuming normality and the compound symmetry constraint (3.1), $100(1-\alpha)\%$ confidence limits L and U for ρ are

$$\begin{aligned} L &= \frac{1+(I-1)\tilde{r}-(1-\tilde{r})F_L}{1+(I-1)\tilde{r}+(I-1)(1-\tilde{r})F_L} \\ U &= \frac{1+(I-1)\tilde{r}-(1-\tilde{r})F_U}{1+(I-1)\tilde{r}+(I-1)(1-\tilde{r})F_U} \end{aligned} \tag{3.8}$$

where $F_L = F_{1-\alpha/2;J-1,(I-1)(J-1)}$ and $F_U = F_{\alpha/2,J-1,(I-1)(J-1)}$, with $F_{p;m,n}$ representing the pth percentile of an $F_{m,n}$ distribution. Confidence limits for λ are $1/(1-L)$ and $1/(1-U)$. The limits (3.8) follow from the fact, established by Kristof (1963), that $f(\rho)/f(\tilde{r}) \sim F_{(J-1),(I-1)(J-1)}$, where $f(x) = (1-x)/[1+(I-1)x]$. Note that the interval (3.8) is different from the interval more commonly given for intraclass correlation (as, for instance, implicitly by Snedecor and Cochran (1989, p. 244)) because it is based on a different experimental design.

4. Two Models and Three Hypotheses

In this section we formulate two different linear mixed models for the blocked design with $K = 1$. In the context of these models we consider three hypotheses, each of which asserts, in a different sense, the absence of a block effect. Both models are

special cases of the general model presented in Section 3.1. Recall that we assumed $\{(Y_{1j}, Y_{2j}, \ldots, Y_{Ij})' : j = 1, \ldots, J\}$ to be iid random vectors. Let

$$E(Y_{ij}) = \mu + \tau_i$$

where $\Sigma\tau_i = 0$, and let

$$\phi_\tau = \frac{1}{I-1}\Sigma\tau_i^2.$$

Thus, ϕ_τ is the usual noncentrality parameter for treatment effects.

As a first step, we decompose the Y_{ij} as follows:

$$Y_{ij} = W_{ij} + e_{ij} \tag{4.1}$$

where the e_{ij} are iid random variables with mean zero and variance σ_e^2. The W_{ij} are random variables that represent the mean value of Y_{ij} that would be obtained from a large number of observations of treatment i in block j, while the e_{ij} represent variation about these means. For instance, in Example 1.3 W_{1j} and W_{2j} would be the actual values of neural density in the right and left hemispheres, while e_{ij} would represent measurement error. We assume that the W_{ij} and the e_{ij} are independent.

4.1 The Models

We now formulate two models for the W_{ij}. (A third model, which unifies Model 1 and Model 2 in a different way than (4.1), was first suggested by Nelder 1954 and developed more fully in Hocking 1985.)

The following model is a special case of Scheffe's (1956, 1959, pp. 261ff) model.

Model 1:

$$W_{ij} = \mu + \tau_i + b_j + g_{ij} \tag{4.2}$$

where the b_j and g_{ij} are random variables which are iid as j varies and for which

$$\begin{aligned}
&\text{(a)}\ \ E(b_j) = 0 \\
&\text{(b)}\ \ \mathrm{Var}(b_j) = \sigma_b^2 \\
&\text{(c)}\ \ E(g_{ij}) = 0 \\
&\text{(d)}\ \ \mathrm{Var}(g_{ij}) = I^{-1}(I-1)\sigma_g^2 \\
&\text{(e)}\ \ \mathrm{Cov}(g_{ij}, g_{i'j}) = -I^{-1}\sigma_g^2, \quad i \neq i' \\
&\text{(f)}\ \ \mathrm{Cov}(b_j, g_{ij}) = 0
\end{aligned} \tag{4.3}$$

Note that (4.3c), (4.3d) and (4.3e) imply

$$\sum_i g_{ij} = 0. \tag{4.4}$$

A slightly different notation for Model 1, preferred by some authors (e.g., Steel and Torrie 1980), is to define σ_g^2 as $\mathrm{Var}(g_{ij})$ rather than as in (4.3d). The disadvantage of this more natural notation is that it necessitates a modification of the usual EMS algorithm.

A second model, given, for example, by Searle (1971, pp. 400–401) is the following.

Model 2:

$$W_{ij} = \mu + \tau_i + \tilde{b}_j + \tilde{g}_{ij} \tag{4.5}$$

where the $\tilde{b}_j$ and $\tilde{g}_{ij}$ are random variables which are iid as j varies and for which

$$\begin{aligned}
&\text{(a)}\ \ E(\tilde{b}_j) = 0 \\
&\text{(b)}\ \ \mathrm{Var}(\tilde{b}_j) = \sigma_{\tilde{b}}^2 \\
&\text{(c)}\ \ E(\tilde{g}_{ij}) = 0 \\
&\text{(d)}\ \ \mathrm{Var}(\tilde{g}_{ij}) = \sigma_{\tilde{g}}^2 \\
&\text{(e)}\ \ \mathrm{Cov}(g_{ij}, g_{i'j}) = 0, \quad i \neq i' \\
&\text{(f)}\ \ \mathrm{Cov}(\tilde{b}_j, \tilde{g}_{ij}) = 0
\end{aligned} \tag{4.6}$$

The key difference between Models 1 and 2 is in assumption (e) of (4.3) and (4.6). In Model 2, the terms $\tilde{g}_{ij}$ and $\tilde{g}_{i'j}$ are uncorrelated while the analogous terms in Model 1 are

negatively correlated. (Additive versions of the models, with g_{ij} and $\tilde{g}_{ij}$ omitted, would be identical.) The covariance condition (4.3e), and its associated constraint (4.4), may at first appear peculiar or arbitrary. Note, however, that because of (4.4) we can write

$$b_j = \overline{W}_{\cdot j} - \mu$$

thus, (4.4) leads to a simple and natural definition of the block main effect b_j as the average over i of W_{ij}, minus its expectation. This natural definition of the main effect b_j is a motivation for (4.4) and, thus, for (4.3e). By contrast, the main effect $\tilde{b}_j$ of Model 2 cannot be so simply defined; we will return to this point in Section 6.2.

For simplicity in subsequent discussion, we now drop the subscript j and write the models as follows:

General:

$$Y_i = W_i + e_i$$

Model 1:

$$W_i = \mu + \tau_i + b + g_i$$

Model 2:

$$W_i = \mu + \tau_i + \tilde{b} + \tilde{g}_i$$

where we assume that the bs, gs satisfy (4.3) and the $\tilde{b}$s and $\tilde{g}$s satisfy (4.6). (Where necessary for clarity, we will revert to the subscripted form of the models.)

4.2 EMS Tables

Table 4.1 shows the EMSs expressed in terms of the model parameters. Note that the EMSs agree with the two versions in Table 2.2 for the case $K = 1$.

Table 4.1 goes here

To clearly locate the source of the discrepancy between the Model 1 and Model 2 EMSs, let us derive directly the EMS(Blocks) for the case $I = 2$. According to Table 3.1,

$$\text{EMS(Blocks)} = \tfrac{1}{2}\text{Var}(Y_1 + Y_2) \tag{4.7}$$

Now, for Model 1,

$$Y_1 + Y_2 = 2\mu + 2b + (e_1 + e_2) \tag{4.8}$$

and for Model 2,

$$Y_1 + Y_2 = 2\mu + 2\tilde{b} + (\tilde{g}_1 + \tilde{g}_2) + (e_1 + e_2) \tag{4.9}$$

Applying the assumptions of each model to (4.8) and (4.9) yields immediately the following EMSs:

$$\begin{aligned} \text{Model 1}: \text{EMS(Blocks)} &= 2\sigma_b^2 + \sigma_e^2 \\ \text{Model 2}: \text{EMS(Blocks)} &= 2\sigma_{\tilde{b}}^2 + \sigma_{\tilde{g}}^2 + \sigma_e^2 \end{aligned} \tag{4.10}$$

Thus, the absence of the g_i terms in (4.8), which is due to the constraint (4.4) imposed by Model 1, leads to the absence of σ_g^2 in (4.10).

Returning to Table 4.1, comparison of the two EMS columns yields the following relations between the model parameters:

$$\begin{aligned} \sigma_b^2 &= \sigma_{\tilde{b}}^2 + I^{-1}\sigma_{\tilde{g}}^2 \\ \sigma_g^2 &= \sigma_{\tilde{g}}^2 \end{aligned} \tag{4.11}$$

(But we will see in Section 6.1 that (4.11) must be taken with a grain of salt.). The relations (4.11) were given by Searle (1971, pp. 403–4, Hocking (1973), and Harville (1978). They provide an algebraic resolution of the paradox of the conflicting EMSs. In the next section we complete the resolution of the paradox by providing a statistical interpretation.

4.3 The Hypotheses

Consider the following hypotheses, each of which in some sense expresses the assertion that there is no difference between the blocks.

$$\begin{aligned} H_1&: \sigma_b^2 = 0 \\ H_2&: \sigma_{\bar{b}}^2 = 0 \\ H_3&: \sigma_b^2 = 0,\ \sigma_g^2 = 0 \\ H_4&: \sigma_{\bar{b}}^2 = 0,\ \sigma_{\bar{g}}^2 = 0 \end{aligned} \tag{4.12}$$

It follows from (4.2)–(4.6) that H_4 is equivalent to H_3. Thus, (4.12) includes three distinct hypotheses: H_1, H_2, and H_3.

We now ask whether (assuming normality of the random variables) the hypotheses in (4.12) can be tested. Consideration of the likelihood function easily confirms what is suggested by Table 4.1 — that neither H_1 nor H_3 can be tested (a consequence of identifiability problems). On the other hand, if the assumptions of Model 2 are satisfied then the statistic MS(Blocks)/MS(T*B) yields a valid F test of H_2.

To interpret the hypotheses H_1, H_2 and H_3, we relate each to the joint distribution of $(W_1, W_2, \ldots, W_I)$. Let

$$\overline{W_\cdot} = \frac{1}{I}\sum_i W_i \tag{4.13}$$

It follows from (4.2), (4.3) and (4.4) that $\overline{W_\cdot} = b$, so that

$$\sigma_b^2 = \mathrm{Var}(\overline{W_\cdot}).$$

Thus, H_1 asserts that $\overline{W_\cdot}$ has variance zero. The stronger hypothesis H_3 asserts that each of the random variables $\{W_i\}$ has variance zero. The following example illustrates these interpretations.

Example 4.1. Recall Example 1.3, where Y is neural density. The hypothesis H_3 asserts that all women in the population have the same neural density W_1 in the left hemisphere, and also that all women in the population have the same neural density W_2 in the right hemisphere. The hypothesis H_1 asserts that the average neural density in the two hemispheres $(\overline{W_\cdot})$ is the same for all women in the population.

To interpret H_2, note from (4.5) and (4.6) that, for $i \neq i'$,

$$\text{Cov}(Y_i, Y_{i'}) = \text{Cov}(W_i, W_{i'}) = \sigma_b^2 \tag{4.14}$$

Thus, Model 2 requires that

$$\text{Cov}(Y_i, Y_{i'}) \geq 0 \tag{4.15}$$

and under this condition the parameter σ_b^2 can be identified with $\rho\sigma_Y^2$ defined in (3.1). Thus, the hypothesis discussed in Section 3, of no efficiency gain due to blocking, expressed as $H_0: \rho = 0$ and tested by the F statistic MS(Blocks)/MS(T*B), can be interpreted within Model 2 as $H_2: \sigma_b^2 = 0$. (Note, however, that within Model 2 the alternative hypothesis must be one–sided, that is, $H_{2A}: \sigma_b^2 > 0$.)

It is worth noting that the F test of H_2 can also be useful in certain situations when the hypothesis of real interest is H_3 (or, equivalently, H_4), which asserts that blocks have no effect whatsoever on the observations Y. Consider the scenario in which the researcher would prefer to ignore blocks entirely. For instance, suppose an experimenter, having run 4 treatments using days as blocks, now wishes to run 2 new treatments and then conveniently overlook the day factor in comparing the 6 treatments. This would be justifiable only if H_3 were true. Since H_3 cannot be tested, and since H_3 implies H_2, the consulting statistician might perform a test of H_2 (on the data for the first 4 treatments); clearly, rejection of H_2 should convince the experimenter of the folly of ignoring days.

In summary, the discussion of this section shows that the conflicting EMS tables express the absence of a "blocks" term in different senses, corresponding to two distinct hypotheses concerning blocks. The hypotheses may be verbally expressed as

H_1: Block main effect is zero

H_2: Blocking has no effect on the
efficiency of treatment comparisons

In the mixed model with interaction present, these hypotheses have different meanings, and only H_2 can be tested if $K = 1$.

5. The Case of Replication ($K > 1$)

The case $K > 1$ represents replicate measurements within the (i,j)th block–treatment combination. For instance, in Example 1.2 one might make K independent measurements on each forearm of each subject. Another example is the generalized randomized blocks design in which each block contains IK experimental units which are randomly allocated to the I treatments.

As in Section 2, we let Y_{ijk} represent the kth observation on the ith treatment in the jth block; we decompose Y_{ijk} as

$$Y_{ijk} = W_{ij} + e_{ijk} \tag{5.1}$$

where the random vectors $(W_{1j},\ldots,W_{Ij})'$ are iid as j varies, and the e_{ijk} are iid random variables with mean 0 and variance σ_e^2 which are uncorrelated with the W_{ij}.

The models for W_{ij} discussed in the preceding section carry over unchanged to the present case. As in the case $K = 1$, there are two seemingly similar hypotheses which actually address different questions. The hypothesis $H_2\colon \sigma_b^2 = 0$ asserts that blocking has no effect on the efficiency of treatment comparisons, whereas the hypothesis $H_1\colon \sigma_b^2 = 0$ asserts that the "true" response W, averaged over treatments, is the same for all blocks.

The entire discussion in Sections 3 and 4 of testing, estimation, and interpretation of σ_b^2 and ρ can be readily carried over to the present case by identifying Y_{ij} of the previous discussion with $\overline{Y}_{ij\cdot}$ of the present case.

In contrast with the previous discussion, as suggested by the Version 1 EMSs in Table 2.2 the hypothesis H_1 can be tested when $K > 1$. The F statistic for this test is the ratio MS(Blocks)/MS(Within T*B), and is, of course, not the same as that used to test H_2.

6. Further Comparison of Model 1 and Model 2

Linear representations like Models 1 and 2 are important heuristic devices in planning and interpreting statistical analyses. In this section we investigate further the contrast between the two models.

6.1 Constraints

First, let us see how each model constrains the joint distribution of $(W_1,\ldots,W_I)$. It follows from (4.2)—(4.6) that both models require the compound symmetry condition

$$\begin{aligned} \mathrm{Var}(W_i) &= \sigma_W^2 \\ \mathrm{Cov}(W_i, W_{i'}) &= \rho_W \sigma_W^2. \end{aligned} \tag{6.1}$$

Moreover, because of (4.14) Model 2 imposes the additional constraint

$$\rho_W \geq 0. \tag{6.2}$$

The constraint (6.2), which is essential to Model 2, is not a trivial one. In fact, it strikes at the heart of Model 1, in the following sense: In Model 1, $\mathrm{Cov}(W_i, W_{i'}) = \sigma_b^2 - I^{-1}\sigma_g^2$, so that the hypothesis $H_1: \sigma_b^2 = 0$ is incompatible with (6.2) unless $\sigma_g^2 = 0$. Consequently, it is impossible to discuss the meaning of H_1 in terms of Model 2. (Confusingly, (4.11) seems to suggest that H_1 is equivalent to $H_4: \sigma_b^2 = 0,\ \sigma_g^2 = 0$; but this is an illusion, because (4.11) cannot hold unless (6.2) holds, and in particular does not hold if $\sigma_b^2 = 0$.)

6.2 Representations

We now ask how a given joint distribution of $(W_1,\ldots,W_I)$ can be represented in Model 1 and in Model 2.

Given any joint distribution satisfying (6.1), it is immediate from (4.2) and (4.3) that a representation in terms of Model 1 can always be found, namely,

$$\begin{aligned} b &= \overline{W_\cdot} - \mu \\ g_i &= W_i - \mu - \tau_i - b \end{aligned} \tag{6.3}$$

where $\mu = E(\overline{W_\cdot})$ and $\tau_i = E(W_i) - \mu$.

For representation by Model 2, the joint distribution of $(W_1,\ldots,W_I)$ must satisfy not only (6.1), but also (6.2). In this case, a representation can be constructed by introducing a random variable Z which is independent of $(W_1,\ldots,W_I)$. The construction is

$$\begin{aligned} \tilde{b} &= c_1\overline{W_\cdot} + c_2 Z + c_3 \\ \tilde{g}_i &= W_i - \mu - \tau_i - \tilde{b} \end{aligned} \tag{6.4}$$

where the constants c_1, c_2 and c_3 depend on the first and second moments of $(W_1, \ldots, W_I)$ and Z.

If Model 1 is additive ($\sigma_g^2 = 0$), then Model 2 is also additive, and in this case the two models coincide and are unique. Otherwise, the construction (6.4) of Model 2 is not unique because the (nondegenerate) random variable Z is arbitrary. Note that any version of Model 2 can be reexpressed in terms of Model 1 through the relations

$$\begin{aligned} b &= \tilde{b} + \bar{\tilde{g}}. \\ g_i &= \tilde{g}_i - \bar{\tilde{g}}. \end{aligned} \tag{6.5}$$

The explicitness of Model 1 contrasts sharply with the indeterminacy of Model 2. Model 1 is constructed in the same spirit as a fixed–effects model, with the random variables W_{ij} playing the role of the population cell means μ_{ij} of the fixed–effects model. The random variables W_{ij} are "in principle observable" in the sense that, if one had enough replication ($K \to \infty$), one could estimate each W_{ij} very closely. (The parameters μ_{ij} in the fixed–effects case are in principle observable in the same sense.) It follows that the parameters μ and τ_i and the random variables b_j and g_{ij} of Model 1 are in principle observable, in the sense that, if one had enough replication and sufficiently many blocks ($J,\ K \to \infty$) one could estimate each term very closely. In particular, the best linear unbiased predictors (BLUPs) (Searle 1987, Sec. 13.4, Harville 1978) of b_j and g_{ij} are consistent, in the sense that, as $J, K \to \infty$ the BLUPs converge (almost surely and in mean square) to the corresponding random variables b_j and g_{ij}.

By contrast, because of the arbitrary random variable Z, if $\sigma_{\tilde{b}}^2 > 0$ and $\sigma_{\tilde{g}}^2 > 0$, then estimation of the random variables $\tilde{b}_j$ and $\tilde{g}_{ij}$ of Model 2 is in principle impossible, no matter how much data one has collected. Although the BLUPs of $\tilde{b}_j$ and $\tilde{g}_{ij}$ converge (almost surely and in mean square) as $J, K \to \infty$, their stochastic limits are <u>not</u> equal to the corresponding random variables $\tilde{b}_j$ and $\tilde{g}_{ij}$; in fact, the expected squared error of prediction converges to $\sigma_{\tilde{b}}^2 \sigma_{\tilde{g}}^2 (I\sigma_{\tilde{b}}^2 + \sigma_{\tilde{g}}^2)^{-1}$. Thus, the BLUPs of $\tilde{b}_j$ and $\tilde{g}_{ij}$ may reasonably be termed inconsistent. (These facts are proved in the Appendix.)

6.3 Interaction versus Independent Contribution

Both Models 1 and 2 have the property of partitioning variance:

$$\text{Model 1:} \quad \text{Var}(Y_i) = \text{Var}(b) + \text{Var}(g_i) + \text{Var}(e)$$

$$\text{Model 2:} \quad \text{Var}(Y_i) = \text{Var}(\tilde{b}) + \text{Var}(\tilde{g}_i) + \text{Var}(e)$$

Nevertheless, Models 1 and 2 have very different heuristic flavors. Model 1 is analytic in flavor in that it decomposes W_{ij} into interpretable components, whereas Model 2 may be termed synthetic, in that it "builds up" W_{ij} from uncorrelated components. Thus, Model 2 has the spirit of a variance components model, whereas Model 1 does not.

Yet in an important sense the interaction term (as expressed within either Model 1 or Model 2) is fundamentally different from a common type of "variance component" term. Very often, a term is added to a model to represent the contribution of a new, independent source of variation. It is common to assume *a priori* that such terms are stochastically independent over all their indices and of the other terms in the model. Let us call such a term an "independent contribution" (IC) term.

The interaction term in the randomized blocks model is not an IC term. Rather, the interaction indicates non-additivity: it is present because W_{ij} cannot be expressed as the sum of a row effect and a column effect. Note, in particular, that if we were to modify our assumptions and consider blocks to be fixed rather than random, then the W_{ij} would be constants (fixed) and consequently the interaction term would necessarily be fixed. By contrast, the randomness of an IC term is inherent; it is not a consequence of the randomness of other terms in the model.

To illustrate, let us consider a more complicated design, the split–plot design. Suppose Y_{ijk} represents the crop yield in the jth block (random) of a randomized block design, on the ith whole–plot treatment (fixed) and the kth subplot treatment (fixed). The usual model for this design is (see Steel and Torrie 1980, p. 393–394):

$$Y_{ijk} = \mu + T_i + B_j + G_{ij} + S_k + R_{ik} + e_{(ij)k} \tag{6.6}$$

where μ, T_i, S_k, and R_{ik} are constants, $\{B_j\}$ are iid random variables, $\{G_{ij}\}$ are iid random variables, and $\{e_{(ij)k}\}$ are iid random variables. Superficially, (6.6) seems to be

an extension of Model 2, with G_{ij} playing the role of g_{ij}. But in fact the G_{ij} term, often termed the whole-plot error, is not simply an interaction term. Note that the G_{ij} term cannot be dropped from the model, even if the T and B effects are believed to be additive; furthermore, G_{ij} would be a random variable even if the T and B effects were regarded as fixed. It is natural to think of G_{ij} as an IC term representing the random contribution of the (ij)th whole plot; with this interpretation the independence of the G_{ij}'s from each other is an assumption about the physical separateness of the plots. (Of course, one might also wish to include $T \times B$ interaction in the model; there would then be two competing models, analogous to our Models 1 and 2, which would place different constraints on the interaction term.)

The preceding discussion suggests that, as a heuristic aid, Model 2 may lead to confusion since it contains a term, $\tilde{g}_{ij}$, which purports to model interaction but closely resembles an IC term.

7. Summary and Discussion

We have considered the blocked design, modeled as a mixed model with the intrablock covariance matrix assumed to be compound symmetric.

7.1 Blocks and Random Factors

In our view, the question of testing blocks is not identical to the question of testing the random factor. The difference arises because there are two distinct null hypotheses concerning blocks, which can be succinctly stated as follows:

H_1: Block main effect is zero

H_2: Blocking has no effect on the efficiency of treatment comparisons

The hypothesis H_1 asserts that the "true" response, averaged over treatments, is the same for all blocks; it is the direct analog of the hypothesis of no main effect in a fixed-effects ANOVA. Assuming normality, H_1 can be tested by the F ratio MS(Blocks)/MS(Within T*B); the test requires within-block replication ($K > 1$).

In a discussion of mixed models, Kempthorne (1975) writes:

> Is there a case for testing the main effect of the random factor? I think it is hard to make one ...

We agree with this statement in reference to H_1, which is what Kempthorne had in mind. Cases like Example 1.3, where H_1 is a natural hypothesis, probably are rather rare. [Such a case may, however, occur in animal genetics, where mixed models are used which incorporate interaction between environment (fixed) and genotype (random); see, for instance, Muir (1985). There appears to be some controversy concerning the choice of models in this context (Fernando et al., 1984; Yamada and Sugimoto, 1988; Ayres and Thomas 1990); because the choice depends on specific genetic considerations, we do not consider it here.]

On the other hand, in the setting of a blocked design, it is often natural to consider H_2, which asserts that blocking has no effect on the variance of inter–treatment contrasts. Assuming normality, H_2 can be tested by the F ratio MS(Blocks)/MS(T*B). (We have also (Sec. 4.3) discussed the usefulness of this F test as a conservative test of the stronger hypothesis H_3 which asserts the absence of block main effects and interactions.)

The hypothesis H_2 can be expressed as

$$H_2: \rho = 0$$

where ρ is the intrablock correlation. The efficiency gain due to blocking is equal to $1/(1-\rho)$. For instance, in the case considered by Kempthorne (1975) of locations (random) and varieties (fixed) of corn, the quantity $1/(1-\rho)$ expresses the effect of blocking by location on the efficiency of comparisons among varieties. (In a field setting, the effect of locations is usually large, but in a greenhouse it may be rather small.) We have also (Sec. 3.8) described a confidence interval for ρ under the normality and compound symmetry assumptions.

Thus, we find that blocks are different from random factors: The hypothesis H_2 and the parameter ρ would normally be of interest only in a blocked design, where the purpose of blocking is to enhance the comparison of treatments.

Many discussions of the randomized blocks design assume that the treatment–block interaction is zero. Such an additivity assumption implies that the conditions of Model 2 are satisfied and that the hypotheses H_1 and H_2 are equivalent. In our opinion, the additivity assumption is unnecessarily restrictive for investigations in which blocks are regarded as random. Indeed, the notion that a treatment can be represented by the average response over a conceptual population of people, plots of ground, or other units is fundamental to statistical thinking. For this reason it appears to us to be useful to develop a view of blocked designs which does not depend on the additivity assumption.

In this connection, we note that lack of additivity often results in lack of compound symmetry. The question of testing and estimating the effect of blocks in the absence of compound symmetry is addressed in a separate paper (Samuels, Casella and McCabe 1990).

7.2 Model 1 and Model 2

We have considered two formulations, Model 1 and Model 2, of the two–way linear mixed model. Model 1 agrees with the two–way mixed model as given in many textbooks; within Model 1, the hypothesis H_1 can be expressed as

$$H_1: \sigma_b^2 = 0.$$

Within Model 2, the hypothesis H_2 can be expressed as

$$H_2: \sigma_b^2 = 0.$$

Thus, each model is linked in a heuristic way with a corresponding hypothesis. (But the F test of H_2 does not require the constraint of nonnegative covariance imposed by Model 2.)

How do our conclusions about testing and estimating blocks reflect on the merits of Models 1 and 2? First, we note that, contrary to the impression conveyed by some textbooks (for instance, Montgomery 1984, p. 222), in most cases the data provide no help in choosing between the models. Except for the Model 2 requirement of nonnegative intra-block covariances, the two models permit exactly the same distributions of the observable

random variables $\{Y_{ijk}\}$. Thus, the data can guide the choice between models only in the limited sense that sufficiently strong negative intrablock covariances will rule out Model 2.

In a real sense, then, Models 1 and 2 are not different models but merely reformulations of the same model, and any choice between them must be made on the basis of their usefulness as heuristic aids. From this viewpoint, both models have drawbacks.

The strengths of Model 1 are its explicitness (Sec. 6.2) and its greater generality. However, Model 1 is heuristically associated with the hypothesis H_1 which is seldom a natural hypothesis for a blocked design.

Model 2 is heuristically associated with the more relevant hypothesis H_2, but, unfortunately, Model 2 has serious conceptual limitations: First, it contains (unnecessarily) random variables which are in principle unobservable and whose BLUPs are inconsistent (see Sec. 6.2). Second, it induces confusion between an interaction term and an independent contribution term (see Sec. 6.3). A third difficulty is that Model 2 constrains intrablock correlations to be nonnegative, a constraint which is irrelevant to testing H_2 and estimating the impact of blocking on efficiency. If the correlation is indeed negative, then blocking decreases, rather than increases, the efficiency of treatment comparisons. This may sometimes happen in practice; for instance, Snedecor and Cochran (1989, p. 243) note that competition between animals in a pen may produce negative intrablock correlations.

An alternative approach, which avoids the need for either Model 1 or Model 2, is for the user to formulate a statistical model directly in term of the covariance structure of the observations (see, for instance, Hocking 1985). We note that many people would find this approach difficult, because for them a linear representation like Model 1 or Model 2 is easier to formulate than a set of assumptions about a covariance matrix.

7.3 Conclusions

We ask, then, which linear representation — Model 1 or Model 2 — would be a more useful aid to straight thinking and to communication between and among statisticians and researchers? We believe that the preponderance of the evidence favors Model 1 as "the"

mixed two–way linear model with interaction. Thus, the Version 1 EMSs displayed in Table 2.2 should be used, rather than the Version 2 EMSs. Confusion concerning the "testing of blocks" in the randomized blocks design could then be avoided by distinguishing clearly between the hypotheses H_1 and H_2, and agreeing that only H_1 is conceptually linked to the EMS table. In this view, the F test for blocks as read from the EMS table is a test of the hypothesis H_1 of no block main effect, whereas the F test of H_2 is regarded as a test of zero correlation, and is unrelated to the EMS table.

Our discussion has been confined to the case of balanced data. If in our setup the number of repeated observations were to vary, say K_{ij} observations on the ith treatment in the jth block, the data would be unbalanced. In the unbalanced case the efficiency gain due to blocking does not bear the simple relationship to the Model 2 parameters that it does in the balanced case. In other respects, the relative merits of Models 1 and 2 are the same in either case, since the models are linear formulations for W_{ij} in the general model $Y_{ijk} = W_{ij} + e_{ijk}$, and the meaning of W_{ij} does not depend on the values of the K_{ij}. Thus, our preference for Model 1 extends to the unbalanced case.

It is true that the distribution theory of the mixed model is more complicated in the unbalanced than in the balanced case, and it is less clear what constitute reasonable estimators of model parameters. But the parameterizations of Models 1 and 2 are linear functions of each other, so any estimation method which is deemed desirable for the estimation of Model 2 parameters would presumably be equally desirable for estimation of the Model 1 parameters.

Our study of the two–way mixed model has led us to believe that in general, in discussing and formulating linear models, it would be useful to distinguish between "interaction" terms and what we have called "independent contribution," or IC, terms. IC terms represent "new" sources of variation and are assumed *a priori* to be stochastically independent over all their indices as well as independent of other terms in the model. An example of an IC term is the whole–plot error terms in the split–plot design. Interaction terms, on the other hand, result from the decomposition of a "true" cell mean. Scheffe (1959, Ch. 8) describes, for a broad class of linear models, the constraints which must

be satisfied by interaction terms defined in this way. In particular, interaction terms in completely crossed designs must sum to zero over the index of each fixed effect.

Since the g_{ij} terms in Model 1 sum to zero, whereas the $\tilde{g}_{ij}$ terms of Model 2 are uncorrelated, the proposed distinction is consistent with our preference for Model 1 as the two–way mixed model with interaction. We emphasize, however, that for variance components investigations, where a goal of the analysis is to estimate components of variance arising from different sources, models similar to Model 2, in that all their random terms are uncorrelated, are often entirely appropriate.

Acknowledgement

We are grateful to the referees for several suggestions which have improved the exposition.

References

Addelman, Sidney (1970), "Variability of Treatments and Experimental Units in the Design and Analysis of Experiments," *Journal of the American Statistical Association*, **65**, 1095–1108.

Anderson, Virgil L. (1970), "Restriction Errors for Linear Models (An Aid to Develop Models for Designed Experiments)," *Biometrics*, **26**, 255–268.

Anderson, Virgil L. and McLean, Robert A. (1974), *Design of Experiments, A Realistic Approach*, New York: Marcel Dekker.

Ayres, Mathew P., and Thomas, Dana L. (1990), "Alternative Formulations of the Mixed–Model ANOVA Applied to Quantitative Genetics," *Evolution* **44**, 221–226.

Cornfield, Jerome and Tukey, John W. (1956), "Average Values of Mean Squares in Factorials," *Annals of Mathematical Statistics*, **27**, 907–949.

Fernando, R.L., Knights, S.A. and Gianola, D. (1984), "On a Method of Estimating the Genetic Correlation between Characters Measured in Different Experimental Units," *Theoretical and Applied Genetics*, **67**, 175–178.

Hartley, H.O. and Searle, S.R. (1969), "A Discontinuity in Mixed Model Analysis," *Biometrics*, **25**, 573–576.

Harville, David A. (1978), "Alternative Formulations and Procedures for the Two–Way Mixed Model," *Biometrics*, **34**, 441–453.

Hicks, Charles R. (1982), *Fundamental Concepts in the Design of Experiments* (3rd ed.), New York: Holt, Rinehart and Winston.

Hocking, R.R. (1973), "A Discussion of the Two–Way Mixed Model," *American Statistician*, **27**, 148–154.

Hocking, R.R. (1985), *The Analysis of Linear Models*, Monterey, California: Brooks/Cole.

Kempthorne, Oscar (1952), *The Design and Analysis of Experiments*, New York: John Wiley.

Kempthorne, Oscar (1975), "Fixed and Mixed Models in the Analysis of Variance," *Biometrics*, **31**, 473–486.

Kirk, Roger E. (1982), *Experimental Design: Procedures for the Behavioral Sciences* (2nd ed.), Monterey, California: Brooks/Cole.

Kristof, Walter (1963), "The Statistical Theory of Stepped–up Reliability Coefficients When a Test Has Been Divided into Several Equivalent Parts," *Psychometrika*, **28**, 221–238.

Kristof, Walter (1972), "On a Statistic Arising in Testing Correlation," *Psychometrika*, **37**, 377–384.

Lentner, Marvin, Arnold, Jesse C. and Hinkelmann, Klaus (1989), "The Efficiency of Blocking: How to Use MS(Blocks)/MS(Error) Correctly," *The American Statistician* **43**, 106–108.

Mehta, J.S. and Gurland, John (1969), "Some Properties and an Application of a Statistic Arising in Testing Correlation," *Annals of Mathematical Statistics*, **40**, 1736–1745.

Montgomery, Douglas C. (1984), *Design and Analysis of Experiments* (2nd ed.), New York:

John Wiley.

Muir, William M. (1985), "Relative Efficiency of Selection for Performance of Birds Housed in Colony Cages Based on Production in Single Bird Cages," *Poultry Science*, **64**, 2239–2247.

Nelder, J.A. (1954), "The Interpretation of Negative Components of Variance," *Biometrika*, **41**, 544–548.

Samuels, Myra L. (1981), "Matching and Design Efficiency in Epidemiological Studies," *Biometrika*, **68**, 577–588.

Samuels, Myra L., Casella, George, and McCabe, George P. (1990), "Evaluating the Efficiency of Blocking Without Assuming Compound Symmetry," submitted to *The American Statistician.*

Scheffe, Henry (1956). "A 'Mixed Model' for the Analysis of Variance," *Annals of Mathematical Statistics* **27**, 23–36.

Scheffe, Henry (1959). *The Analysis of Variance*, New York: John Wiley.

Searle, S.R. (1971a), "Topics in Variance Component Estimation," *Biometrics*, **27**, 1–76.

Searle, S.R. (1971b), *Linear Models*, New York: John Wiley.

Searle, S.R. (1987), *Linear Models for Unbalanced Data*, New York: John Wiley.

Snedecor, George W. and Cochran, William G. (1989). *Statistical Methods* (8th ed.), Ames, Iowa: The Iowa State University Press.

Steel, Robert G.D. and Torrie, James H. (1980), *Principles and Procedures of Statistics* (2nd ed.), New York: McGraw–Hill.

White, Robert F. (1975), "Randomization and the Analysis of Variance," *Biometrics* **31**, 555–572.

Wilk, M.B. and Kempthorne, O. (1955), "Fixed, Mixed, and Random Models," *Journal of the American Statistical Association*, **50**, 1144–1167.

Wilk, M.B. and Kempthorne, O. (1956), "Some Aspects of the Analysis of Factorial Experiments in a Completely Randomized Design," *Annals of Mathematical Statistics*, **27**, 950–985.

Winer, B.J. (1971), *Statistical Principles in Experimental Design* (2nd ed.), New York: McGraw–Hill.

Yamada, Y., Itoh, Y. and Sugimoto, I. (1988), "Parametric Relationships between Genotype × Environment Interaction and Genetic Correlation when Two Environments Are Involved," *Theoretical and Applied Genetics*, **76**, 850–854.

APPENDIX. Asymptotic Behavior of BLUPs for Models 1 and 2

The BLUPs for both Model 1 and Model 2 were given by Harville (1978). We now show that the BLUPs have the asymptotic properties claimed in Section 6.2.

Model 1

Assume that the data can be represented by Model 1, and let b'_j and g'_{ij} denote the BLUPs of b_j and g_{ij}, respectively.

Proposition 1. As $J, K \to \infty$ (in either order) the BLUPs b'_j and g'_{ij} converge almost surely and in mean square to b_j and g_{ij}, respectively.

Proof. From Harville (1978, pp. 446–447), b'_j can be written as

$$b'_j = \frac{IK\sigma_b^2}{IK\sigma_b^2 + \sigma_e^2}(\overline{Y}_{\cdot j \cdot} - \overline{Y}_{\cdots})$$

From this and (5.1), the strong law of large numbers implies that, as $K \to \infty$,

$$b'_j \to \overline{W}_{\cdot j} - \overline{W}_{\cdot\cdot} \tag{A1}$$

(all convergences of random variables are almost sure and in mean square). From (4.2) we have

$$\overline{W}_{\cdot j} = \mu + b_j. \tag{A2}$$

Another application of the strong law shows that as $J \to \infty$, $\overline{W}_{\cdot\cdot} \to \mu$, so that, from (A1) and (A2), $b'_j \to b_j$. The proof that $g'_{ij} \to g_{ij}$ is analogous. Similar arguments establish the same limits if $J \to \infty$ before $K \to \infty$.

Model 2

Assume that the data can be represented by Model 2, and let $\tilde{b}'_j$ and $\tilde{g}'_{ij}$ represent the BLUPs of $\tilde{b}_j$ and $\tilde{g}_{ij}$, respectively.

Proposition 2. As $J, K \to \infty$ (in either order) the BLUPs $\tilde{b}'_j$ and $\tilde{g}'_{ij}$ converge almost surely and in mean square to random variables b^*_j and g^*_{ij} such that

(a) $$b^*_j + g^*_{ij} = \tilde{b}_j + \tilde{g}_{ij}$$

(b) $$E(b^*_j - \tilde{b}_j)^2 = \frac{\sigma_{\tilde{b}}^2 \sigma_{\tilde{g}}^2}{I\sigma_{\tilde{b}}^2 + \sigma_{\tilde{g}}^2}$$

Proof. Harville (1978, pp. 446–447) gives $\tilde{b}'_j$ as

$$\tilde{b}'_j = \frac{IK\sigma_b^2}{IK\sigma_b^2 + K\sigma_g^2 + \sigma_e^2}(\overline{Y}_{\cdot j\cdot} - \overline{Y}_{\cdots})$$

Reasoning analogous to the proof of Proposition 1, using (4.5) instead of (4.1), establishes that $\tilde{b}'_j \to b^*_j$ as $J, K \to \infty$, with

$$b^*_j = \kappa(\tilde{b}_j + \overline{\tilde{g}}_{\cdot j}) \qquad (A3)$$

where $\kappa = I\sigma_b^2(I\sigma_b^2 + \sigma_g^2)^{-1}$. A similar derivation, using $\tilde{g}'_{ij}$ as given by Harville (1978), shows that

$$\tilde{g}'_{ij} \to g^*_{ij} = W_{ij} - \mu - \tau_i - b^*_j. \qquad (A4)$$

The proposition follows easily from (A3) and (A4) and the fact that $\tilde{b}_j$ and $\tilde{g}_{\cdot j}$ are independent.

Table 2.1. Orthogonal Decomposition of the Total Sum of Squares

Source	df	SS
Treatments	$I-1$	$JK\sum_i(\overline{Y}_{i..}-\overline{Y}_{...})^2$
Blocks	$J-1$	$IK\sum_j(\overline{Y}_{.j.}-\overline{Y}_{...})^2$
T*B	$(I-1)(J-1)$	$K\sum_i\sum_j(\overline{Y}_{ij.}-\overline{Y}_{i..}-\overline{Y}_{.j.}+\overline{Y}_{...})^2$
Within T*B	$IJ(K-1)$	$\sum_i\sum_j\sum_k(Y_{ijk}-\overline{Y}_{ij.})^2$
Total	$IJK-1$	$\sum_i\sum_j\sum_k(Y_{ijk}-\overline{Y}_{...})^2$

Table 2.2: Two Versions of Expected Mean Squares for Mixed Model with Treatments Fixed, Blocks Random

	Expected Mean Squares	
	Version 1	Version 2
Treatments	$JK\phi_T + K\sigma_G^2 + \sigma_e^2$	$JK\phi_T + K\sigma_G^2 + \sigma_e^2$
Blocks	$IK\sigma_B^2 + \sigma_e^2$	$IK\sigma_B^2 + K\sigma_G^2 + \sigma_e^2$
T*B	$K\sigma_G^2 + \sigma_e^2$	$K\sigma_G^2 + \sigma_e^2$
Within T*B	σ_e^2	σ_e^2

Table 3.1. ANOVA Table for $I = 2$ and $K = 1$

Source	SS	df	EMS
Treatments	$\frac{1}{2}J(\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot})^2$	1	$J\phi_\tau + \frac{1}{2}\mathrm{Var}(Y_1 - Y_2)$
Blocks	$2\sum_j(\overline{Y}_{\cdot j} - \overline{Y}_{\cdot\cdot})^2$	$J-1$	$\frac{1}{2}\mathrm{Var}(Y_1 + Y_2)$
T*B	$\frac{1}{2}\sum_j[(Y_{1j} - Y_{2j}) - (\overline{Y}_{1\cdot} - \overline{Y}_{2\cdot})]^2$	$J-1$	$\frac{1}{2}\mathrm{Var}(Y_1 - Y_2)$

Table 4.1. EMSs for Model 1 and Model 2 when $K = 1$

	EMS	
Source	Model 1	Model 2
Treatments	$J\phi_\tau + \sigma_g^2 + \sigma_e^2$	$J\phi_\tau + \sigma_{\bar{g}}^2 + \sigma_e^2$
Blocks	$I\sigma_b^2 + \sigma_e^2$	$I\sigma_b^2 + \sigma_{\bar{g}}^2 + \sigma_e^2$
T*B	$\sigma_g^2 + \sigma_e^2$	$\sigma_{\bar{g}}^2 + \sigma_e^2$